



dr hab. inż. Marcel Młyńczak

Politechnika Warszawska

Wydział Mechatroniki

Instytut Metrologii i Inżynierii Biomedycznej

Recenzja rozprawy doktorskiej

mgr. inż. Jakuba Piskozuba

„A data glove with a reduced number of sensors for recognition of Polish Sign Language Letters”

Dziedzina: nauki inżynieryjno-techniczne

Dyscyplina: inżynieria biomedyczna

Promotor: prof. dr hab. inż. Paweł Strumiłło

Podstawa opracowania recenzji

Recenzja została przygotowana w odpowiedzi na pismo prof. dr hab. inż. Ewy Piętki, Przewodniczącej Rady Dyscypliny Inżynieria Biomedyczna, RDIB.002.122.2023, z dnia 28.11.2023 r. Ocenie podlegało spełnienie przez rozprawę warunków określonych w art. 187 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce.

Ocena układu rozprawy doktorskiej

Rozprawa doktorska została napisana w języku angielskim i zawiera 66 stron tekstu głównego (nie licząc strony tytułowej, streszczenia, abstraktu i spisu treści, które jednak nie zostały ponumerowane). Składa się ze wstępu, wprowadzenia teoretyczno-technologicznego, opisu projektów rękawic sensorowych, metodologii badawczej, wyników, dyskusji, wniosków, bibliografii i krótkiego suplementu. Jej układ jest w związku z tym standardowy dla prac naukowych. W pracy pojawia się lista skrótów, jednak nie zawarto zwyczajowo dodawanych również list grafik i tabel. Streszczenie i abstrakt są relatywnie krótkie, przez co nie zawierają podstawowych szczegółów metodologii oraz uzyskanych wyników. Niestandardowy jest także sposób wprowadzenia skrótów, poprzez zastosowanie myślników, również na końcu zdania (przykład: „- PSLA -”).

Omówienie treści i ocena rozprawy doktorskiej

Wstęp (rozdział 1) zawiera cel i motywację pracy, tezę oraz przedstawienie struktury pracy (w przypadku ostatniej sekcji powtarza się błędnie numeracja sekcji „1.1”). Celem pracy było opracowanie rękawicy sensorowej wspomaganej metodami uczenia maszynowego do oceny możliwości jej wykorzystania

w diagnostyce i rehabilitacji (wątek nie został w pracy szerzej rozwinięty) w kontekście rozpoznawania liter polskiego alfabetu, z minimalną liczbą sensorów. W opisie motywacji pracy powołano się m.in. na to, że dotychczasowe rozwiązania mają niedostateczną ergonomię i niezadowalającą skuteczność. Określona została teza pracy – możliwe jest efektywne rozpoznanie liter polskiego alfabetu migowego z wykorzystaniem rękawicy sensorowej z ograniczoną liczbą sensorów. Na tym etapie nie określono jednak, co rozumie się przez ograniczoną ich liczbę, a w dalszej części pracy nie wskazano również jednoznacznie kryteriów, które będą brane pod uwagę (wspomniane w tezie uproszczenie konstrukcji i poprawa ergonomii nie musi zawsze iść w parze z ograniczeniem liczby sensorów).

Wprowadzenie teoretyczno-technologiczne (rozdział 2) stanowi przegląd składni i alfabetu polskiego języka migowego (zdaje się, że Rysunek 1.1.2 ma niepoprawny numer – powinno być 2.1.2); punktu widzenia potencjalnych użytkowników (zauważono m.in. problem z brakiem bezpośredniej informacji o mimice twarzy, którą wykorzystuje się w trakcie komunikacji z wykorzystaniem języka migowego); przeglądu opisanych w literaturze rozwiązań (rysunek 2.3.2 został wykorzystany w pracy Autora pracy, jednak nie przytoczono odpowiedniej referencji), spośród których większość stanowią rozwiązania oparte o sensory wizyjne, a wśród rękawic sensorowych – rozwiązania wykorzystujące czujniki piezorezystancyjne, optyczne, jak również MEMS; następnie przeglądu metod uczenia maszynowego wykorzystywanych jako człon analizujący dane zbierane z sensorów (wspomniana metoda DTW jest zazwyczaj stosowana jako element wspomagający i poprzedzający modelowanie, nie będący bezpośrednio metodą uczenia maszynowego); ciekawego opisu rozwiązań rękawic sensorowych ze zwróceniem uwagi na problemy oraz uzyskiwane dokładności; na koniec przedstawiony został spis podstawowych metod redukcji wymiarowości („dimensionality reduction”) oraz oceny ważności cech („feature selection”).

Dwie pierwsze wprowadzające części zostały opisane na 28 stronach, co stanowi ponad 42% całej pracy.

Opis autorskich projektów rękawic sensorowych przedstawiony został w rozdziale 3. W jego pierwszej części zawarte zostało wyjaśnienie konkretnych wyborów technologicznych. W pierwszym kroku Autor zdecydował się na wykorzystanie czujników piezorezystancyjnych. Wyznaczona została charakterystyka zmian rezystancji w zależności od kąta zgięcia. Obserwując zbliżoną do liniowej zależność, ustalono, że surowe wartości otrzymywane z sensora nie będą konwertowane. Wydaje się, że jest to zasadne przybliżenie, jednak z metodologicznego punktu widzenia wspomnianą charakterystykę powinno się zweryfikować dla kilku przypadków i nie łączyć punktów pomiarowych łamaną.

W kolejnej części przedstawiono trzy wersje rękawicy. Pierwsza składała się z 5 czujników piezorezystancyjnych. Zwrócona została uwaga na problem z wpływem przewodów na rejestrację sygnałów. W drugiej wersji zwiększono liczbę czujników do 10, jednocześnie poprawiając ich układ i montaż. Wyniki w rozpoznawaniu gestów okazały się niesatysfakcjonujące. Wydaje się jednak, że zdanie „Unsatisfactory results in classifying the gestures tested with it revealed the need to completely change the sensor mounting system.” powinno zostać bardziej rozwinięte. Trzecia wersja rękawicy składała się z tej samej liczby i układu czujników jak wersja druga, z tą różnicą, że czujniki zostały inaczej przyklejone do rękawicy, a ścieżka ruchu czujnika była stabilizowana za pomocą wydrukowanych w 3D prowadnic, które były przymocowane do rękawicy za pomocą rzepów.

Przedstawiony został również moduł pomiarowy oraz kontrolny, w obecnej wersji oparty o mikrokontroler Broadcom BCM2835. Rysunek 3.3.1 stanowi adaptację rysunku z publikacji Autora, ponownie bez przywołania odpowiedniej referencji.

Pomimo istotnego wkładu pracy na etapie poprawek wprowadzanych do kolejnej wersji, działania te miały charakter typowo inżynierski, a nie naukowo-badawczy. Widać to zwłaszcza w części dotyczącej opracowania transmisji bezprzewodowej, którą należy postrzegać jednak jako ciekawą opcję techniczną, bez bezpośredniego wpływu na cel i tezę pracy.

Wskazanie oraz ocena zastosowanych metod badawczych

Rozdział 4 zawiera opis metodologii. Zaplanowano 4 etapy ewaluacji. W pierwszym etapie, nazwanym „Pilot”, wykorzystano pierwszą wersję rękawicy. Oceniono sensowność zastosowanych czujników, ich liczbę i rozmieszczenie. Nie wskazano jednak dokładnie, jak dokonano tej oceny. Na tym etapie nie były również rejestrowane dane u ludzi. W drugim etapie („Evaluation”) oceniano „aplikowalność” rękawicy (nie wskazano jednak jednoznacznie której) w rozpoznawaniu statycznych gestów w grupie 3 osób.

W trzecim etapie (tym razem już bez nazwy) ponownie oceniano możliwość klasyfikacji gestów statycznych, jednak na większym zbiorze danych pochodzącym z rejestracji w grupie 15 osób, z wykorzystaniem metod uczenia maszynowego. Na tej podstawie możliwa była również ocena ważności cech, a w konsekwencji również wskazanie najistotniejszych dla tego zadania sensorów (docelowo w celu ograniczenia ich liczby). W ostatnim, czwartym etapie, wykorzystano podejścia oparte o rekurencyjne uczenie głębokie do oceny gestów dynamicznych w grupie 16 osób.

Podstawową uwagą na tym etapie, która nie została również przedyskutowana jako ograniczenie pracy, jest oparcie badań na rejestracjach relatywnie małej grupy badanej, zwłaszcza pod kątem wyciągania wniosków z analizy z wykorzystaniem metod głębokiego uczenia. Nie jest do końca prawdą, że próbki są niezależne (jak wskazano na końcu str. 38), a fakt 10-krotnego powtórzenia czynności powinien zostać uwzględniony w protokole w taki sposób (fragment nie dotyczy analizy „spersonalizowanej”), że do zbioru treningowego trafiają wszystkie rejestracje danej osoby (w przeciwnym wypadku istnieje bardzo duże ryzyko nieuprawnionej poprawy wyników, w związku z założeniem, że poszczególne rejestracje tej samej osoby są do siebie podobne).

W sekcji dotyczącej grup badanych przedstawiono kryteria wyboru osób do badania, które zakładały brak zaburzeń osteomotorycznych dłoni, brak limitu wieku oraz płci, oraz posiadanie ręki w określonym rozmiarze pasującym na badaną rękawicę. W trzecim etapie, gdzie badane były jedynie gesty statyczne, stwierdzenie (str. 37), że żaden z badanych nie był poddany ekspozycji na język migowy przed badaniem, można postrzegać jako znaczące ograniczenie. Wydaje się jednak, że w czwartym etapie, podczas którego parametryzacji podlegał proces dynamiczny poruszania dłonią, stwierdzenie, że tylko jedna osoba знаła język migowy, jest bardzo poważnym błędem metodologicznym. W pracy nie przedstawiono również opisu procesu uczenia badanych do korzystania z alfabetu migowego.

W opisie protokołu pomiarowego wskazano, że ocenie podczas etapu trzeciego podlegało 16 liter statycznych, natomiast podczas czwartego etapu – wszystkie 36 liter, zarówno statycznych, jak i dynamicznych. Ponownie rysunki 4.2.1 oraz 4.2.2 nie zostały opatrzone referencją z publikacji Autora.

Na uwagę zasługuje uzyskanie zgody komisji etycznej Politechniki Łódzkiej. Nie wspomniano jednak jednoznacznie, czy badani podpisywali formularz świadomej zgody do udziału w badaniu i wykorzystania danych do celów naukowych. Wątek ten pojawia się w suplemencie, jednak w podanym linku do repozytorium Github nie znaleziono „templates of the form informing the subjects about the experiments” (dostęp 28 grudnia 2023 r. oraz 24 stycznia 2024 r.).

W sekcji dotyczącej wstępnego przetwarzania danych, zwrócono uwagę na występowanie artefaktów i obserwacji odstających. Wydaje się, że ten wątek powinien zostać dokładniej opisany statystycznie. Jest to jeden z potencjalnych aspektów utrudniających zastosowanie metody w praktyce. Dla czytelnika nie jest również jasne, jakie cechy zostały (w przypadku klasyfikacji statycznej) wyznaczone.

Sekcja dotycząca metod uczenia maszynowego zawiera listę metod wykorzystywanych na etapie drugim, trzecim i czwartym. Jednak należy podkreślić, że więcej szczegółów przedstawiono jedynie dla metod głębokiego uczenia, co utrudnia replikowalność badań przeprowadzonych dla liter statycznych.

Ostatnia sekcja przedstawia listę 4 metod wykorzystywanych do oceny ważności cech. Brakuje w niej jednak szczegółów bądź referencji. Jako błąd należy również wskazać, że rysunki 4.6.1 i 4.6.2 przedstawiają już wyniki, a nie metodykę (jednak bez dokładniejszego opisu pozostają nieczytelne dla czytelnika).

Podsumowując, opis metodologii jest niepełny, przedstawia procedurę analizy stosowaną raczej w pracach inżynierskich (oraz czasem magisterskich), w mojej ocenie brakuje w nim aspektu badawczego. Wydaje się także, że trudno wyciągać wnioski z rejestracji na tak mało licznej grupie badanych, spośród których wszyscy, oprócz jednej osoby, nie znali wcześniej języka migowego. Realną ocenę przetwarzania danych oraz wykorzystywanych metod uczenia maszynowego uniemożliwia również brak wspomnianych w suplemencie danych oraz skryptów Python (dostęp 28 grudnia 2023 r. oraz 24 stycznia 2024 r.).

Ocena części rozprawy doktorskiej dotyczącej omówienia wyników badań

Rozdział 5 przedstawia uzyskane wyniki dla opisanych w części metodologicznej etapów 2-4.

Podczas etapu drugiego badane zostały 3 osoby, przedstawiono jednak dla 3 metod uczenia maszynowego (tabela 5.1.1) jedynie wyniki bez wskazania, jak dokonano podziału na zbiór treningowy i testowy, czy wykorzystywano podejście „cross-validation”. Jako drobną uwagę edycyjną należy również wskazać stosowanie do oddzielenia części dziesiętnej przecinków zamiast kropek (inna konwencja w języku angielskim w porównaniu do języka polskiego), a także niezastosowanie identycznej liczby miejsc po przecinku w ramach jednej kolumny tabeli. Podano, że szerszy opis wyników został zamieszczony w pracy Autora [56]. Jeśli jednak praca doktorska nie stanowi „zszywki” publikacji, powinno się w niej zamieścić podstawowe szczegóły, których w rozprawie zabrakło.

Sekcja 5.2 zawiera przegląd wyników dokładności w klasyfikacji liter statycznych. Ponownie nie zostało przedstawione (właściwie powinno to znaleźć się w części z opisem metodologii), jak dokonany został podział na zbiór treningowy i testowy. Co ciekawe, na początku rozdziału prezentującego wyniki podano definicję czułości i specyficzności, jednak później jedyną stosowaną miarą była dokładność (pozostałe

pojawiają się jedynie w tekście na str. 51). Ponownie odwołano się do faktu, że szerszy opis wyników został zamieszczony w pracy Autora [57].

Sekcje 5.3 oraz 5.4 zawierają przegląd wyników klasyfikacji z wykorzystaniem metod rekurencyjnych. Tym razem przedstawiono metodykę podziału danych na trzy podzbiory.

Ciekawe porównania wyników z innymi danymi literaturowy Autor przedstawił w sekcji 5.5. Jednak ze względu na to, że protokoły, sensory i wykorzystywane metody analityczne były różne, należy je postrzegać jedynie poglądowo. Stąd tego typu fragmenty czasem pojawiają się dopiero w sekcji prezentującej dyskusję pracy (zwłaszcza, że w dyskusji nie zawarto już bezpośrednich odniesień do innych prac).

Informacje dotyczące praktycznego zastosowania uzyskanych wyników badań

Rozdział 6 obejmuje dyskusję oraz konkluzje. W pierwszej sekcji podsumowane zostały wyniki. Zauważono, że do tej pory żaden zespół naukowy podejmujący podobny problem naukowy nie analizował istotności poszczególnych lokalizacji, a co za tym idzie - możliwości zmniejszenia liczby sensorów.

Druga sekcja stanowi zebranie najistotniejszego wkładu, jaki praca wnosi do dziedziny rozpoznawania liter alfabetu migowego z wykorzystaniem rękawicy sensorowej z ograniczoną liczbą sensorów. Autor zwrócił uwagę na aspekt przewagi podejścia opartego o spersonalizowane modele uczone na bazie zbioru danych po augmentacji. Autor pracy zaznaczył, że zaprezentowane wyniki mogą posłużyć zmniejszeniu rozmiaru rękawiczek, które mogą zostać częścią systemów tłumaczenia języka migowego. Aby tak jednak było, należałoby uzupełnić opis o wiele szczegółów technicznych i metodologicznych (wspomnianych wcześniej), jak również przeprowadzić badania w grupie osób faktycznie posługujących się językiem migowym w warunkach zbliżonych do naturalnego funkcjonowania.

W tym fragmencie pojawia się sugestia, że intrygujące okazało się uzyskanie lepszych rezultatów dla zbioru po augmentacji przeprowadzonej dla jednej osoby w stosunku do sytuacji uczenia na danych zebranych w grupie badanych. Efekt ten jest dość oczywisty i łatwy do wyjaśnienia. Różnym metodom uczenia maszynowego „łatwiej” jest zbudować „model” danych wejściowych, jeśli są one bardziej spójne.

Następnie pojawia się komentarz, że przeprowadzenie dużej liczby pomiarów wydaje się nie być wymagane do utworzenia uniwersalnego modelu, a krótka kalibracja przeprowadzona na jednej osobie wydaje się wystarczająca. Trudno wyobrazić sobie, jak w takiej sytuacji uzyskać można by odpowiednią generalizację macierzy wejściowej. Podejścia spersonalizowane tworzy się zazwyczaj w taki sposób, aby dane konkretnej osoby „dodawać” do wstępnego dużego zbioru danych na etapie „re-treningu” modelu. Taki komentarz został nawet dodany przez Autora w kolejnym akapicie, ale już w innym kontekście.

Analiza potencjalnych ograniczeń jest niezwykle krótka i skupia się jedynie na aspektach technologicznych. Sekcja dotycząca perspektyw przyszłych badań jest interesująca, aczkolwiek należy ją postrzegać jako poglądową.

Ocena, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego

Dyskusja nie zawiera zwyczajowo dodawanej sekcji podsumowującej oryginalny wkład autora do dziedziny. Antycypując jej zawartość, można by wskazać: ograniczenie liczby sensorów, zastosowanie rękawicy do rozpoznawania relatywnie dużego zbioru liter polskiego języka migowego (składających się z 16 liter statycznych i 20 dynamicznych), czy zastosowanie augmentacji na etapie przygotowania danych do trenowania modeli uczenia maszynowego. W mojej ocenie zarówno pierwsza, jak i trzecia kwestia, ma charakter inżynierskiego usprawnienia, a nie oryginalnego rozwiązania problemu naukowego. Druga kwestia ma znów raczej charakter nowatorski, a nie oryginalny.

Ocena zastosowanego piśmiennictwa

Bibliografia pracy zawiera relatywnie niewielką (jak na pracę doktorską) liczbę 59 pozycji, w tym 14 stron internetowych (bez podania daty dostępu), 4 prac konferencyjnych oraz 2 prac typu „pre-print” opublikowanych na portalu arxiv.org. Dwie referencje są przywołane w sposób niepełny: [49] oraz [59].

Autor powołał się także na dwie prace własne – pozycje [56] i [57]. Nie jest jednak jasne, dlaczego w pierwszej spośród tych pozycji nie zostali wypisani wszyscy współautorzy (na stronie internetowej oraz w publikacji - <https://pe.org.pl/articles/2019/9/29.pdf> - autorów jest czterech).

Źródła są przeważnie nowe, zróżnicowane i adekwatne do treści pracy. Potwierdzają zapoznanie się Autora z dotychczasowymi osiągnięciami w dziedzinie.

Ocena, czy rozprawa prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej

Przedstawiona do recenzji rozprawa prezentuje wiedzę teoretyczną Autora w zakresie konstrukcji rękawic sensorowych, wprowadzania usprawnień technologicznych, czy wykorzystania metod uczenia maszynowego. Nie ma też bezpośrednich podstaw, aby wątpić, że Autor potrafi samodzielnie prowadzić pracę naukową. Uwzględniając jednak liczne uwagi zebrane w poprzednich częściach recenzji należy podkreślić, że praca w sposób niewystarczający (a, w mojej ocenie, w niektórych fragmentach niepoprawny) przedstawia aspekt badawczy przeprowadzonych prac - możliwy do oceny zwłaszcza w sekcjach opisujących metodologię, wyniki oraz ich dyskusję.

Autor opublikował dwie prace jako pierwszy autor. Pierwsza, w języku polskim, w czasopiśmie „Przegląd elektrotechniczny” (IF = 0,5; MNiSW = 70). Druga, w języku angielskim, w czasopiśmie „Applied Sciences” (IF = 2,7; MNiSW = 100). Odwołując się do wymagań, w przypadku publikacji wieloautorских powinny zostać załączone oświadczenia o indywidualnym wkładzie procentowym i merytorycznym wszystkich współautorów. Nie zostały one jednak dołączone do zestawu podlegającego recenzji.

Autor występował również z pracą „Data Glove for the Recognition of the Letters of the Polish Sign Language Alphabet” na konferencji Polish Conference on Biocybernetics and Biomedical Engineering 2023 (DOI:10.1007/978-3-031-38430-1_27).

Na stronie Google Scholar oraz na portalu ReaserchGate Autor ma do tej pory (25 stycznia 2024 r.) zarejestrowane 3 cytowania (w tym jedno autocytowanie). Baza Scopus wykazuje 2 cytowania.

Konkluzja końcowa

W mojej ocenie, biorąc pod uwagę wcześniej wyrażone opinie i uwagi, zakres i poziom naukowy uzyskanych wyników nie odpowiada w pełni ustawowym i zwyczajowym wymaganiom stawianym rozprawom doktorskim określonym w ustawie Prawo o szkolnictwie wyższym i nauce z dn. 20 lipca 2018 r. W związku z powyższym wnioskuję do Wysokiej Komisji powołanej przez Radę Dyscypliny Inżynieria Biomedyczna Politechniki Śląskiej o nieprzyjęcie rozprawy i niedopuszczenie mgr. inż. Jakuba Piskozuba do jej publicznej obrony.

Warszawa, 25.01.2024 r.

Marcel Młyńczak

